Method for enhancing text entry rate with single-switch scanning

Heidi Horstmann Koester, PhD;[1*] Richard Callaghan Simpson, PhD, ATP[2]
[1]*Koester Performance Research, Ann Arbor, MI;* [2]*Department of Electrical and Computer Engineering, New York Institute of Technology, Old Westbury, NY*

Abstract—Single-switch scanning is a technique used by some individuals for spoken and/or written communication. We developed a method for adjusting the settings in a single-switch scanning interface to increase a user's text entry rate (TER). We evaluated that method with nine individuals who use single-switch scanning to communicate. Text entry rates improved by an average of 120% ($p = 0.003$). All nine subjects increased their TER by at least 40%, and five of the nine increased their TER by over 100%. At baseline, TER averaged 1.42 words per minute (wpm), ranging from 0.28 to 2.92 wpm. With the revised settings, TER averaged 2.72 wpm and ranged from 1.12 to 6.51 wpm.

Key words: alternative communication, assistive device, assistive technology, augmentative communication, computer access, physical impairment, single-switch scanning, switch access, text entry, user performance.

INTRODUCTION

Single-switch scanning allows people with severe physical impairments, who may also be unable to speak, to independently use a computer or augmentative and alternative communication (AAC) device. The impairments may stem from a variety of medical conditions, such as cerebral palsy, traumatic brain injury, muscular dystrophy, and neuromuscular diseases such as multiple sclerosis and amyotrophic lateral sclerosis, affecting hundreds of thousands of people in the United States. Physicist Stephen Hawking is perhaps the most well-known user of single-switch scanning.

While single-switch scanning affords spoken and written communication using only one controlled input movement, it is a slow method of text entry. A very fast user may achieve 7 or 8 words per minute (wpm) [1–4], while rates of 1 wpm and lower are common [5–8]. Despite its limitations, scanning is often the only alternative for individuals who cannot use other interfaces. Technologies such as eye gaze and speech recognition may not be usable for individuals with severe spasticity, poor head control, or limited verbal abilities. Direct brain interfaces, while promising, are still in the development stage [9].

The most common implementation of scanning is row-column scanning, which can be used with as little as one switch for input. Typical operation with one switch requires two switch hits to select an item from a two-dimensional matrix of letters, numbers, symbols, words, or phrases. Each row of the matrix, beginning with the first, is highlighted in turn until the first switch hit is made to select a row. Each column of the row is then highlighted in turn until the target is highlighted, when the second switch hit is made to select the target. Variations on this theme include group-row-column scanning,

which adds another level in which a group is first selected, followed by the rows then columns in that group, as well as manual initiation, which requires an additional switch hit to resume scanning after a selection is made. **Figure 1** shows an example of a group-row-column layout used by one of this study's participants.

Given the challenges of efficient text entry with just a single switch, product developers have implemented numerous features and configuration settings to allow for customization of scanning software with the goal of increasing text entry rate (TER). Some of the settings available in current scanning systems are shown in **Table 1**.

In addition to these settings, another major aspect of configuration is the layout and position of the items in the selection matrix. Each item has specific scan distance, defined by the number of scan periods required to reach it. In a row-column matrix, items in the upper left corner have the shortest scan distance, while those in the lower right have the highest. Placing the most frequently used items in the locations with the shortest scan distance is a common strategy for enhancing TER [10–12].

Proper configuration of the features available within scanning systems can make a major difference in communication rate [10,13–14]. The timing parameters are a
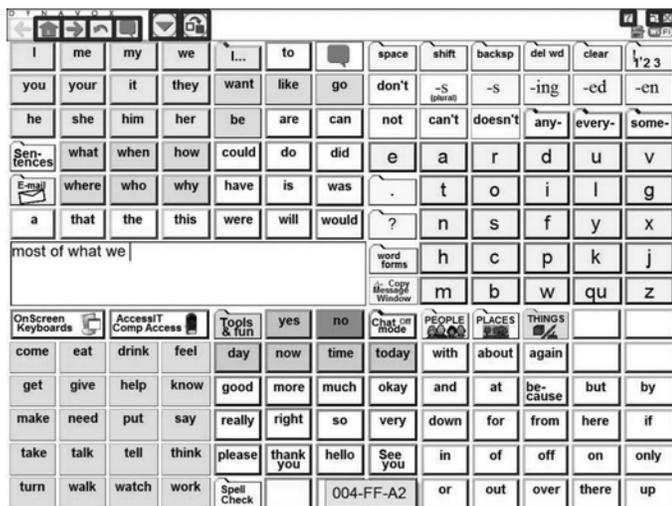
Figure 1.
Example of group-row-column scanning display (from subject 004's baseline system). It includes 8 groups, scanned in following order: (1) titlebar, (2) message window, (3) pronouns, (4) helping verbs, (5) letters+prediction, (6) verbs, (7) chat, (8) prepositions. Once a group is selected, scanning continues through rows, and then columns, of that group.

key factor in TER, particularly the scan delay, which defines the amount of time an item is available for user selection. As a simple example, if a person is capable of using a scan delay of 1.0 s, but their system is set to 2.0 s, their TER will be roughly half of what it could be. One case study demonstrated how modifications to both item layout and scan delay yielded an increase in TER of 321 percent for one individual [11], and the five individuals in another study showed differences of 20 to 25 percent when using different configurations [6].

Our goal is to establish an effective and efficient process for tailoring a scanning interface to a particular user. As an early step in this process, we conducted this study to define and evaluate a procedure that helps identify the most appropriate scanning settings for a given individual. Demonstrating the success of the method will provide the basis for further development of software tools to make it easier and more efficient for practitioners to optimize scanning interfaces for single-switch scanning users.

RESEARCH GOALS AND HYPOTHESES

The goal of this study was to evaluate the effectiveness of a method for enhancing performance of single-switch scanning users. Specifically, we hypothesized that the revised settings resulting from our method would significantly improve users' TER as compared with baseline. We did not have a specific hypothesis regarding the effect on errors across all subjects since this was expected to vary depending on baseline performance. For individuals whose baseline included very few errors, we did not expect errors to significantly improve; for those who had many errors at baseline, we expected the revised settings to reduce errors.

METHODS FOR ADJUSTING SCANNING SETTINGS

Our method for enhancing a user's single-switch scanning interface is shown in flowchart form in **Figures 2** and **3** and is summarized in **Table 2**. The current process is designed to be carried out by a skilled practitioner and is supported by software measurement tools that assess each user's abilities and help identify the effect of adjusting various configuration parameters. The key elements of the process are described here.

Table 1.
Configuration options found in commercially available scanning interfaces [5].

Category	Configuration Option	Definition
Language Features	Character Prediction	One or more items in matrix are dynamically updated based on which letters are most likely to be selected next.
	Word Prediction	One or more items are dynamically updated based on what word user is most likely entering or is likely to enter next. Additional settings control number of words in prediction list, when list is displayed, etc.
	Fixed Words	One or more items in matrix contain fixed words or phrases.
Item Positions	Group Layout	How groups of items are positioned relative to each other.
	Item Layout	Positions of letters or other items relative to each other.
Scan Pattern	Number of Levels	Nesting of levels in scan pattern (e.g., group-row-column or row-column).
	Manual/Automatic Initiation	Manual requires switch hit to initiate scanning. Automatic resumes scanning automatically after selection.
	Loop Count	Number of scans through columns in row before returning to row scan.
	Start from Last Selection	Whether to resume scanning from top after selection, or resume from position of selected item.
	Up-a-Level Item	Exits current scan and restarts one level up.
Dead Time	Scan Message Window	Include message display window as item in scan pattern.
	Scan Title Bar	Include title bar as item in scan pattern.
	Postselection Delay	Time delay after each selection before scanning resumes.
Timing	Scan Delay	Length of time item is highlighted and available for selection.
	1st-Item Delay	Delay added to provide extra scanning time for first row or column.
	Repeat Delay	Repeats selection if switch is held down longer than repeat delay.
	Acceptance Delay	Length of time switch must be activated before activation is registered. Reduces bounce (unintentionally pressing twice).

Switch Test

The first step in the process is to measure the user's ability to reliably activate the switch and identify any problems with the current switch site or activation method. In our study, we used the Switch test within our Compass assessment software (Koester Performance Research; Ann Arbor, Michigan) that asks the user to complete a series of single-, double-, and triple-presses. The Compass Switch test measures the time it takes the user to press the switch in response to a visual prompt. Note that this is a more straightforward task than using the switch to make a scanning selection because it requires less precise timing and anticipation. The average switch-press time and the frequency of switch-press errors (specifically errors from "bouncing" on the switch) are reported. The test also recommends an appropriate scan delay and recovery delay to use during scanning. These recommendations are based on the user's switch-press times and have been validated in previous work [15].
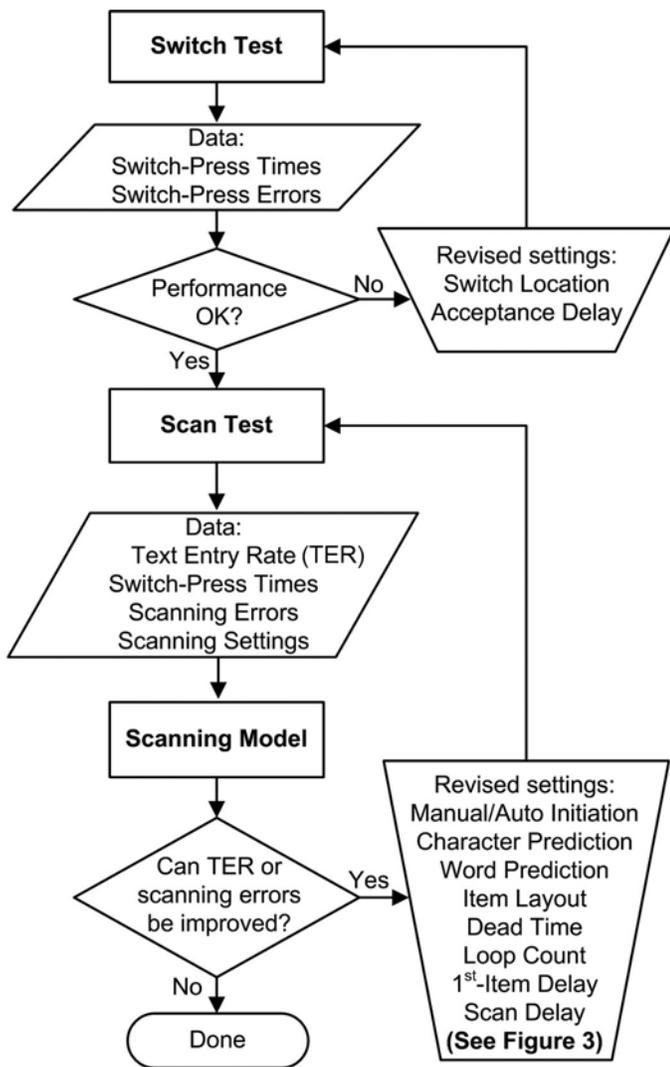
Scan Test

Once isolated performance with the switch is consistent, the next step in the process is to assess the user's ability to use the switch to make scanning selections. In our study, we examined this in two ways. First, each participant transcribed one sentence within the Compass Scan test, which presents a letter matrix for text entry then automatically counts scanning errors and measures TER. Second, we asked users to enter two sentences using their current scanning system. We video-recorded the user's screen and, when available, logged each selection made using the language activity monitoring features built into some AAC devices. These data were analyzed manually to identify scanning errors and to calculate baseline TER against which to gauge any improvements.

Scanning Modeling Software

To support the process of identifying interventions, we used our Scanning Model Software (SMS), a program that simulates the TER achieved with a particular single-switch scanning configuration [5,16]. SMS takes as input the configuration of a letters-only scanning interface and the rates of eight types of scanning errors and outputs the TER that would be expected under those conditions. The configuration is input by specifying the matrix of items, the timing parameters, the scanning pattern, and the user's

Figure 2.
Method for revising settings of single-switch scanning interface.

Figure 3.
Decision tree for identifying revised settings. Init = initiation, N = no, Y = yes.

cation. The single-switch scanning users in this study all used a variety of linguistic features such as word prediction and whole-word items, but all relied on single-letter spelling for a large part of their text entry. We used SMS to help optimize letter-by-letter spelling for the users in this study, but its use is not required to apply our method for identifying interventions.

Identifying Interventions

Data from the Switch and Scan tests, as well as simulations from the Scanning Model component, are used to identify specific modifications to the user's scanning interface. **Figure 3** illustrates the general approach to this. As a rough rule of thumb, if the user's scanning errors exceed 25 percent of his or her correct selections, modifications are first made to reduce errors. A "scanning error" occurs for any timing error during scanning and any time the user selects an incorrect item. Examples of timing errors include letting the system scan through all the rows without selecting the desired row or selecting the row just after the desired row.

Once the error rate has been reduced, additional modifications are made to increase overall efficiency. This is supported by our SMS that helps simulate the effect of modifications before they are made. For example,

switch-press times. The time required to select a given item is the sum of the time required to scan to the item and press the switch the required number of times. Our model also included the time penalty imposed by each type of error and error correction method, along with the likelihood of each error occurring. For example, if the user failed to select the target row the first time it was highlighted, the system would scan through all the rows in the matrix once and then scan through the rows again until it reached the target row.

The output of the model is the TER for the letter layout specified in the input. This is a simplification because it supports only single-letter items, but a useful simplifi-

Table 2.
Method for modifying configuration of single-switch scanning interface.

Conceptual Step	Procedure
1. Identify problems with current switch activation.	Run single-, double-, and triple-hit Switch tests in Compass.
	Observe errors or difficulties in activating switch, and note number of extra hits reported by Compass test.
	Revise switch type or location if switch activation is not consistent.
2. Gauge adequacy of current scan delay.	Use results from Switch test and its recommended scan delay and 1st-item delay. (Note: recommended scan delay is calculated as single-hit switch-press time multiplied by 1.5.)
3. Identify problems with using switch to make scanning selections.	Run Compass Scan test that mimics current letter layout and settings.
	Note error counts in resulting report.
	Complementary approach is to ask user to enter text with their scanning system. Record video or make careful observations to count and classify scanning errors.
4. Accommodate any problems identified in step 3.	Error reduction is first focus if scanning errors exceed 25% of correct selections. **Figure 3** illustrates some changes that can be considered, such as slower scan delay, manual initiation, additional dead time, or clearer scan pattern.
5. Assess success of accommodations in step 4.	Remeasure scanning errors as in step 3. Ideally, get scanning errors down below 25%.
6. When scanning errors are roughly below 25%, identify enhancements for efficiency.	First, establish fastest timing parameters that allow low-error use. These will come from step 2 but may have been adjusted during error reduction step 4.
	Then, try to reduce scan steps at minimal extra cost. Consider the following, where applicable:
	<u>Group organization:</u>
	–Put letters+prediction group first.
	–Stay on letters+prediction group until word has been completed.
	–Ensure that other groups serve useful purpose; else remove them.
	<u>Language features:</u>
	–Remove character prediction.
	–Use 3 to 6 words in list.
	–Provide enough time to search list, e.g., with either postselection time or manual initiation.
	<u>Letter layout:</u> use frequency-based layout, unless strong reason to do otherwise.
	<u>Scan pattern:</u>
	–Make sure it's sensible, with orderly progression.
	–Consider manual initiation only if errors are problem or layout is unfamiliar.
	–Set loop count to 1; higher loop counts are generally not good idea.
	<u>Dead time:</u>
	–Identify any sources of dead time (titlebar, message window, postselection delay).
	–If exists, is there good reason for it?
	–If not, eliminate.
	<u>Scanning Model Software software simulates effect of these, but is not mandatory:</u>
	–Timing parameters.
	–Letter layout.
	–Manual/automatic initiation.
	–Loop count.
	–Reduction in scanning errors.

three types of adjustments considered by our method are shown in **Figure 3** and described subsequently.

"Dead time" refers to time between the last selection being made and the resumption of scanning for the next selection. There is no dead time if a system is configured to immediately and automatically start scanning a language item after the previous selection is made. However, if a system scans one or more items prior to scanning language items or introduces a fixed delay after an item is selected, then dead time is present and might contribute to a slower TER. One way to reduce dead time is to change the scanning order so that frequently used language groups are scanned first in the scanning cycle ahead of functional items such as the window title bar, which are only needed occasionally.

Inefficient use of word prediction means that the user is not deriving enough keystroke savings to compensate for the additional scan steps introduced by word prediction. This "inefficiency" may be the result of a poorly configured word prediction system (i.e., an inappropriate lexicon, too few or too many words in the prediction list), or it may be the result of a suboptimal usage strategy on the part of the user. The solution depends on the exact cause of the problem, but a common remedy is to ensure that the user has sufficient time to search the prediction list.

Revising the item layout is another revision that can enhance efficiency. The item layout defines the positions of each item in the scanning grid. An efficient layout presents the most frequently used items before less frequently used items. For example, a scanning grid with letters laid out in frequency-based order is 25 percent more efficient than a layout in alphabetical order [10].

METHODS FOR EVALUATION STUDY

Study Design

This study used a longitudinal ABA design to evaluate the effectiveness of our method for enhancing performance of single-switch scanning users. We compared the performance of single-switch scanning users with the settings that they use every day (condition A, baseline settings) to their performance with the settings identified with our method (condition B, revised settings).

Subjects

Participants were recruited through United Cerebral Palsy of Pittsburgh, Pennsylvania, and with the assistance of a speech-language pathologist in Ann Arbor, Michigan. The recruitment plan and experimental protocol were approved by the University of Pittsburgh's Institutional Review Board. All participants provided informed consent. Subjects qualified for the study if they had used single-switch scanning for at least 6 mo and were able to transcribe text from written copy using their scanning system.

Ten subjects were enrolled, and nine subjects completed the protocol. (One user was unable to complete the text transcription task because of literacy issues.) All nine subjects regularly used single-switch scanning to access an AAC system. Seven used their system for "all conversations," while two used it for "most" because they could mouth some words successfully. All used letter-by-letter spelling combined with some type of word prediction; six individuals also had fixed single-word items available. One subject (003) used auditory scanning because of a visual impairment. Six subjects activated their switch with head motion; three used hand activation. All had used their system for at least 1 yr, and typically around 5 yr, prior to entering the study. Subjects' devices were from five different manufacturers. Self-reported satisfaction with these devices averaged 3.9, with a standard deviation of 1.4, on a scale of 1 (very dissatisfied) to 5 (very satisfied). Eight subjects had the help of a practitioner to establish their initial configuration and reported changing their settings about once a year. One subject had no help, professional or otherwise, and used the default settings on his system. A summary of the subjects is provided in **Table 3**.

Baseline Phase (Condition A1)

In the first session of the study, each subject's baseline performance was measured on his or her original (preintervention) scanning system and configuration. Each completed two transcription tasks, of two sentences each, as follows:

1. *Freeform* text entry, in which users entered text using their own system in whatever way they saw fit. They were free to use any prediction items, single-word items, etc., that were present on their system, at their own discretion.

2. *Letters-only* text entry, in which users entered text using letter-by-letter spelling only. This used the same settings as in the freeform test but restricted the user to a specially created letters-only test page, in the same layout normally used by the subject.

Table 3.
Key characteristics of participants and their baseline systems.

Subject	Sex	Age (yr)	Diagnosis	System	Letter Layout	Prediction	Scan Rate (s)	Satisfaction
001	F	53	CP	Dynavox DV4	Frequency-based	2-character, 4-word	1.5	3
002	F	41	CP	PRC Vantage, Word-Power	QWERTY	6-word	0.9	5
003	M	21	CP	Dynavox Vmax	Alphabetic	6-word	2.6	5
004	M	17	CP	Dynavox Maestro, Scanning Word-Power	Frequency-based	6-character, 4-word	1.5	4
006	F	19	CP	Dynavox Vmax, Picture Word-Power	QWERTY	16-word	1.0	3
007	F	28	CP	Words+ EZKeys	Frequency-based	6-word	0.9	5
008	F	37	CVA	Viking Communicator 4, Scanning WordPower	Frequency-based	3-word	0.8	4
009	M	50	CP	Words+Say-It-Sam	QWERTY	8-word	1.3	1
010	F	54	ALS	Viking Communicator 4	Frequency-based	8-word	1.0	5

ALS = amyotrophic lateral sclerosis, CP = cerebral palsy, CVA = cerebrovascular accident, F = female, M = male.

Subjects were allowed to make errors and decide whether or not to fix them [17]. Subjects also completed the Compass Switch and Scan tests using their switch attached to the investigator's laptop.

Identify Intervention

The method described previously was used to identify revised settings for each subject. The revised settings were entered into each subject's scanning system and any layout changes were implemented to use for the remainder of the study. All decisions and revisions were made by the research team.

Intervention Phase (Condition B)

In four weekly sessions, each subject completed a two-sentence freeform test and a two-sentence letters-only test. The order of conditions was randomized and counterbalanced across subjects. The text used in each test was unique, but all sentences were equivalent in terms of letter frequency and reading level. Subjects were given 4 wk with the revised settings to provide time to get used to the new configuration. All subjects except one (002) used the revised settings during daily life between sessions. Subject 002 did not like the revised settings at first, so she reverted to her baseline settings between sessions.

Reversal Phase (Condition A2)

After 4 wk of using the revised settings, each subject's baseline settings were restored to his or her scanning system. As in the other phases, subjects completed a freeform and a letters-only test.

Survey Questions

At the end of the study, subjects were asked to complete the following questionnaire, using a Likert scale anchored with 1 = strongly disagree and 5 = strongly agree:

1. Overall, I now prefer the new settings to my original ones.
2. I did not like the new settings at first.
3. I think I typed faster with the new settings.
4. I would like to keep some of the new settings to use permanently.

Dependent Variables

Text Entry Rate

TER was measured for each test as the number of correct characters present by the completion of the test divided by the total time for the test. The total time

included all time from start to end of transcription, including any time for fixing incorrect selections. This characters per second measure was converted to words per minute by assuming five characters per word. TER was calculated for each freeform and letters-only transcription test.

Scanning Errors

A scanning error was counted any time the user selected an incorrect item and for any timing error during scanning. We recorded the total number of scanning errors in each transcription test. Total scanning errors included the following error types:

- Failing to press the switch when the target group, row, or column was highlighted.
- Pressing the switch too early (before the intended target).
- Pressing the switch too late (after the intended target).

Selection errors are a subset of the total scanning errors that resulted in the wrong item (i.e., column) being selected. Total selection errors included all selection errors in the session, while net selection errors were the number of selection errors that were not corrected by the end of the task. All errors were reported as a percentage of the correct selections made in the test.

Data Analysis

Paired *t*-tests (with an alpha of 0.05) were conducted to examine the main effect of baseline versus revised settings for the following variables:

- TER (correct words per minute).
- Total scanning errors (%).
- Total selection errors (%).
- Net selection errors (%).

In each statistical test, baseline (condition A) was calculated from the average of session A1 and A2 results for each subject and revised (condition B) was the results from each subject's fourth session with the revised settings (B4). Additionally, data from each individual subject were examined for a clear increase in TER with revised settings relative to baseline and a full reversal back to baseline performance when settings were reverted to their original values. Full reversals add confidence that any significance in the group statistics was truly due to enhancement provided by the revised settings rather than extraneous or random factors.

Responses to the questionnaire were analyzed using Nielsen's guidelines for 5-point Likert scales [18]. Responses to positive questions (Q1, Q3, and Q4) were considered significant if the mean response to that question was greater than 3.6 based on a one-sample *t*-test at a $p = 0.05$ level. The response to the negative question (Q2) was considered significant if the mean response was less than 2.4 at a $p = 0.05$ level. Using a stricter standard than the neutral value of 3.0 accommodates for subjects' tendency to be polite when responding to these types of questions.

RESULTS

Characteristics of Modified Configurations

As shown in **Table 4**, at least one change was made for each subject. These are described further subsequently, with specific details provided for subject 004 as one concrete example (see subject 004's baseline configuration in **Figure 1** and revised configuration in **Figure 4**).

Switch Configuration

For one subject, we changed the position of the switch from lying flat on the front middle of the wheelchair

Table 4.
Changes made to each subject's scanning system with revised settings versus baseline. Check-mark indicates that setting in this category was changed in revised settings relative to original.

Settings Category	Subject								
	001	002	003	004	006	007	008	009	010
Language Features	✓	✓	—	✓	✓	—	—	—	—
Item Positions	✓	✓	✓	✓	✓	—	✓	✓	✓
Scan Pattern	—	✓	—	✓	✓	—	✓	✓	—
Dead Time	✓	✓	—	✓	✓	—	✓	—	—
Timing Parameters	—	✓−	✓+	✓+	✓−	✓+	✓+	✓−	✓+
Switch	—	—	—	—	✓	—	—	—	—
✓+ = faster speed with revised, ✓− = Slower speed for revised.									

Figure 4.
Revised configuration for subject 004. (See **Figure 1** for base-line.) Scanning now begins with letters+prediction group in upper-left section.

laptray to sitting perpendicular to the laptray in the front right corner. Changing the position also changed the activation site from the hands (the subject used both) to the right elbow. An activation delay was also introduced to limit the number of unintentional switch activations.

Timing Parameters

Scan delay was increased for four subjects and decreased for three subjects. The largest changes were subject 002 increasing from 0.9 to 1.6 s and subject 003 decreasing from 2.6 to 2.1 s. Recovery delay was reduced for two subjects and increased for one subject. For subject 004, scan delay stayed at 1.5 s, but the baseline recovery delay of 0.27 s was removed.

Dead Time

All the systems we encountered in this study could be configured to scan one or more items prior to scanning language items, which introduces dead time. For five subjects, we removed the titlebar and/or message window from the top of the scan tree, either by moving them farther down in the progression or replacing them with commands directly in the main scanning matrix. In subject 004's case, both the titlebar and message window were removed from the scan pattern. Note, however, that occasionally we kept the titlebar and/or message window at the top of the scan tree to provide time for the user to

search the word prediction list. We want the right amount of dead time: not too much, not too little.

An additional source of dead time was a postselection delay, which was a fixed delay after an item (a column) was selected. During the postselection delay, a user can repeat the selection by hitting the switch again (e.g., to get double letters). It also gives the user a chance to check the word list before scanning resumes from the top. A postselection delay was removed from the original configuration of one subject (005). This individual already scanned the message window to allow time for searching the prediction list, and she did not need the extra postselection delay.

Item Positions

We made at least one item position change to eight of the nine subjects' configurations. For four subjects, we rearranged groups of items to simplify or optimize the scanning pattern. In subject 004's case (as well as subject 008), we moved the letters+prediction group to the first position. For seven subjects, we changed their letter layout to a more optimal frequency-based arrangement; this replaced a QWERTY layout for three subjects and enhanced the existing frequency arrangement for four others. For subject 004, the letters were moved up by two rows to reduce the scanning distance to them.

Scan Pattern

The overall flow of scanning was modified for three subjects. For two of these, including subject 004, we adjusted the system so that when the letters+prediction group was selected, scanning remained on that group until the word was completed. For the third, the overall flow was simplified from a group-row-column pattern to a row-column pattern; this simplification was also done for the letters+prediction group of subjects 004 and 008. The method of initiating scanning was changed from manual (i.e., the subject had to press the switch to initiate scanning) to automatic (i.e., scanning began immediately after a selection was made) for one individual. We reduced the loop count for two subjects, decreasing the number of times the system would scan through the items within a row without a selection being made. One subject's system was configured at baseline to resume scanning from the point where the previous selection was made. We changed this to the more traditional "start from the top."

Language Features

Language features consisted of character prediction, word prediction/completion, and items with fixed individual words (typically high-frequency words like *the*, *and*, *I*). For the two subjects who were using character prediction, including subject 004, we entered text with and without character prediction using their systems and found that the scan distance was more efficient without character prediction than with it. We therefore removed it from those subjects' configurations. For similar reasons, we did not add character prediction to any subject's configuration. Note that the poor performance of character prediction was surprising given the benefits found in simulations by Lesher et al. [10]; it may be that character prediction performance is sensitive to its particular implementation.

We did not add word prediction to any subject's configuration, nor did we remove it from any subject's configuration where it was in use. In two cases, we did reduce the number of items in the prediction list. We removed individual single-word items from one subject's configuration after observing that she almost never used them.

Baseline Versus Revised Settings for Freeform Text Entry

The main comparison was for freeform text entry since it allowed participants to use all features of their system, however they chose, to generate text. The freeform baseline condition represents the performance afforded by subjects' current system as configured prior to study enrollment. Overall results of our statistical analyses are shown in **Table 5**. TER for each subject in each trial is shown in **Table 6**. TER with the revised settings was significantly better than baseline ($p = 0.003$). Revised settings were at least 40 percent faster than baseline for all subjects, and the average difference over all subjects was 119.7 percent. As shown in **Figure 5**, five of nine participants had TER improve by at least 100 percent. Further, improvement in TER was not limited to the participants with the slowest baseline TER. Of the two fastest subjects at baseline, one improved by over 100 percent, and the other improved by almost 50 percent. Finally, as shown in **Figure 6**, the criterion for full reversal was also met, strengthening the conclusion that the revised settings truly enhanced TER substantially. These statistics demonstrate the enhanced performance

Table 5.
Statistical analyses for freeform text entry.

| Variable	Mean		Difference (%)	*p*-Value
	Baseline	**Revised**		
TER (wpm)	1.42	2.72	119.7	**0.003**
Total Scanning Errors (%)	59.07	15.13	43.94	0.09
Total Selection Errors (%)	10.02	2.99	7.02	0.21
Net Selection Errors (%)	4.83	1.68	3.15	0.34

Note: Bold indicates statistically significant.
TER = text entry rate, wpm = words per minute.

Table 6.
Text entry rate for freeform text entry (words per minute).

Subject	Session					
	A1	**B1**	**B2**	**B3**	**B4**	**A2**
001	1.27	1.41	1.41	1.86	2.12	1.51
002	0.75	0.89	0.97	1.00	1.30	0.71
003	0.55	0.74	0.92	0.82	1.12	0.54
004	1.23	2.74	2.78	2.21	3.23	1.50
006	0.54	1.13	0.94	0.89	1.50	0.57
007	2.18	2.19	2.38	2.44	3.38	1.93
008	3.02	3.16	4.17	4.22	6.51	2.82
009	0.29	0.91	0.70	0.87	1.15	0.28
010	2.84	3.61	3.27	3.11	4.17	2.95
Mean	1.41	1.86	1.95	1.94	2.72	1.42
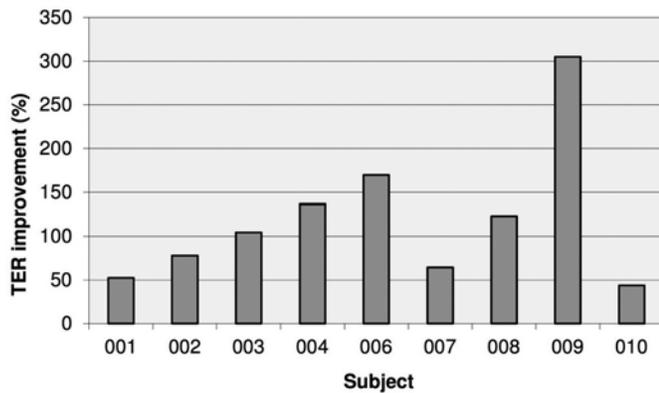
Figure 5.
Increase in freeform text entry rate (TER) for each subject.
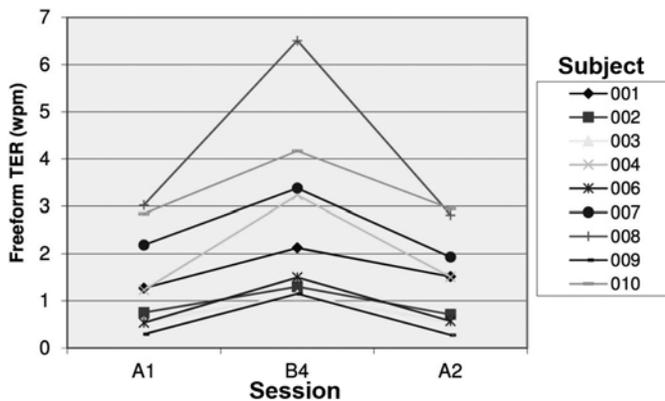
Figure 6.
Baseline (A1), intervention (B4), and reversal (A2) text entry rate (TER) for each subject, showing symmetrical reversal pattern. wpm = words per minute.

achieved after 4 wk with the revised settings; TER also increased steadily during the 4 wk, starting with a 33 percent average improvement in the first week.

The total scanning errors committed by each subject are shown in **Table 7**. No significant effect of settings was found. However, scanning errors averaged 59 percent of selections during baseline text entry but only 15 percent with the revised settings ($p = 0.09$). As shown in **Figure 7**, this difference is largely due to error reductions for four subjects: 002, 003, 006, and 009. This was a big part of the success of revised settings for these individuals.

Comparing Freeform and Letters-Only Text Entry

A comparison of freeform and letters-only text entry is a secondary question, but still of interest. Results of

TER analysis are shown in **Table 8**. In the baseline condition, freeform text entry had no real advantage over letters-only text entry. With the revised settings, however, freeform did provide a significant advantage compared with letters-only (80% faster). The freeform feature used most often by subjects was the word prediction list.

Questionnaire Results

Subject responses to the poststudy questionnaire are shown in **Table 9**. Subjects significantly preferred the new settings ($p = 0.02$) and expressed a significant desire to keep the new settings permanently ($p < 0.001$). Subject responses were not quite significantly above the 3.6 benchmark regarding whether they felt the new settings improved their typing speed ($p = 0.09$, 95% confidence interval 3.5–4.7) and were neutral about not liking the new settings at first ($p = 0.47$). Some loved them right away and some disliked them immediately. All subjects did keep their revised settings active in their system for continued use after the study.

DISCUSSION

Effect of Modified Settings on Performance

The method we used to define modified settings for these subjects' scanning system appears to be quite effective. Our results strongly support the hypothesis that the revised settings would significantly improve TER. As compared with baseline, TER with the revised settings in condition B more than doubled. Five of nine subjects improved their TER by over 100 percent, and everyone achieved at least a 40 percent improvement. Even participants who were already quite effective with their baseline settings improved their TER significantly.

Results are also consistent with our hypothesis that the revised settings would not necessarily affect error rates across all subjects. Five participants already had very low scanning error rates at baseline, around 10 percent or lower, and they stayed at those low rates once they were familiar with the revised settings. Four participants had baseline error rates above 50 percent; for two of these individuals, Subjects 006 and 009, error rates were above 100 percent, meaning that there were more errors than correct selections. The revised settings led to dramatic decreases in error rates for these individuals, down to an average error rate below 25 percent.

Table 7.
Total scanning error rate with freeform text entry (as percentage of correct selections).

Subject	Session					
	A1	B1	B2	B3	B4	A2
001	8.1	27.8	19.3	7.1	8.3	9.1
002	70.0	37.8	18.6	24.0	18.6	89.4
003	54.0	19.7	29.5	30.4	23.3	50.9
004	11.5	9.6	0.0	10.8	2.5	14.0
006	130.4	20.6	40.6	36.4	24.5	140.0
007	6.7	24.1	20.3	19.0	11.4	18.3
008	0.0	57.1	6.0	7.8	8.1	4.3
009	232.5	83.3	83.0	56.1	33.3	216.2
010	1.8	1.9	14.3	14.0	6.1	6.0
Mean	57.2	31.3	25.7	22.9	15.1	60.9

There were two general factors that account for the success of the revised settings. First, reduction in total scanning errors was a key for four subjects (002, 003, 006, and 009). As shown in **Figure 7**, these subjects had very high error rates with their baseline settings, averaging more than one error for every correct item selection. With revisions such as a slower scan delay or more straightforward scan pattern (see **Table 4** for each subject's revisions), errors declined dramatically, and TER increased as a result. The second main factor was increased efficiency, which benefited all subjects. A faster scan delay, decreased dead time, and revised item
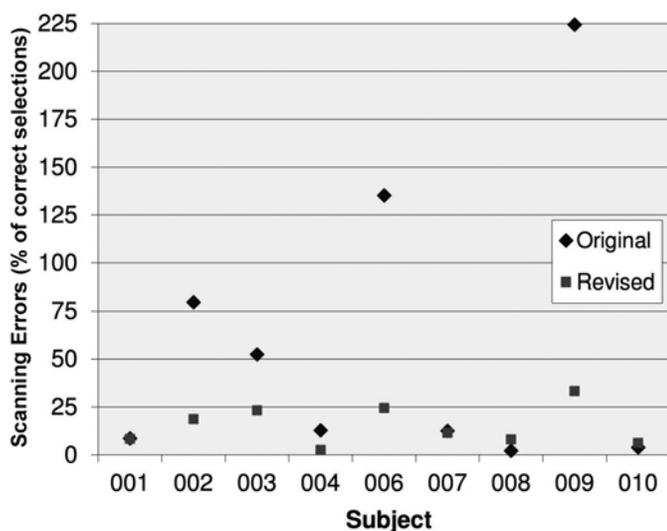
Figure 7.
Total scanning error rates during freeform text entry for each subject, using baseline and revised settings.

positions are examples of changes that can increase efficiency if done properly.

Effect of Freeform Versus Letters-Only Text Entry

These results also provide some insight into how performance using more complex linguistic features compares with simpler letters-only text entry, although this was a secondary question for this study. Freeform text-entry rates were significantly higher than letters-only rates only when revised settings were used. In fact, freeform rates with the revised settings averaged 80 percent higher than letters-only. Yet at baseline, freeform and letters-only were essentially equal. Because the freeform condition allowed subjects to use the word prediction and other linguistic features on their system, these results strongly suggest that the revised settings allowed for more effective use of these features compared with baseline. The advantage of freeform with the revised settings contrasts sharply with other reports in the literature that have found little to no effect of word prediction on TER with scanning [2]. This appears to be a case where the details make the difference; it may be that word prediction can be effective at enhancing TER for single-switch scanners, but only if the system is configured appropriately.

Putting These Results into Context

Surprisingly, this study is the only published report we have seen of TERs achieved by experienced single-switch scanners, using their own system in their own way, as in the baseline freeform condition. TERs in baseline freeform for these nine individuals ranged from 0.28 to 2.92 wpm.

The fact that TERs were consistently below 3 wpm at baseline is significant given early research that suggested

Table 8.
Comparison of freeform vs letters-only text entry rate (TER).

Variable	Mean TER (wpm)		Difference (%)	*p*-Value
	Freeform	Letters-Only		
TER, Baseline	1.42	1.28	10.4	0.27
TER, Revised	2.72	1.51	80.1	**0.01**

Note: Bold indicates statistically significant.
wpm = words per minute.

Table 9.
Responses to poststudy questionnaire.

Subject	Question			
	Q1. Now Prefer	Q2. Initial Dislike	Q3. Faster	Q4. Permanent
001	4	3	3	5
002	4	5	3	5
003	5	2	4	5
004	5	3	5	5
006	4	2	5	5
007	3	1	4	4
008	4	1	4	5
009	5	5	5	5
010	4	3	4	5
Mean	4.22	2.78	4.11	4.89

3 wpm as a minimally acceptable rate for productive communication [19]. At rates below this point, Good-enough-Trepagnier et al. found that conversations break down, primarily due to the receiver's impatience [19]. None of the nine participants averaged higher than 3 wpm at baseline with either freeform or letters-only text entry. With the revised settings, four of nine subjects reached this level. While this is significant progress, it would be ideal to find ways that enabled everyone to reach this minimum threshold.

These rates are perhaps lower than might have been expected given the literature. It is common to assume that single-switch scanning provides a TER of 6 wpm or above [20–21]. In fact, only one of our participants achieved a rate above 6 wpm (subject 008, at 6.51 wpm), even with the revised settings. A closer look at the literature shows that it is mostly based on subjects who have typical motor function. Across seven group studies using unimpaired subjects, the group mean for single-switch scanning ranged from 2.5 to 8 wpm, with TER typically reaching at least 6 wpm with sufficient practice [2–4,7,20–22].

In contrast, the literature involving single-switch users with motor impairments comprises two individual case studies and three group studies, and in all cases, participants used an experimental system that was not their own [4–6,11,15]. The rates in these studies range from about 0.5 to 4 wpm, aligning more closely to those observed here.

Clinical Implications

The results of this study highlight two important principles for helping individuals get the most out of their single-switch scanning systems. First, measuring performance is a critical complement to observation and user feedback. Second, carefully configuring the system is absolutely necessary. Application of these principles is described here.

The relatively low mean and range of TER achieved by these subjects, especially at baseline, highlight the importance of measurements as a key indicator of how well an AAC system is actually working for an individual. The fundamental measurements to take include—

1. Switch hit time as a determiner of the best scan delay for the individual and an indicator of how well the current switch site is working for the user.
2. Scanning errors as an indicator of how consistently and accurately the individual can use their switch to make scanning selections. A total scanning error rate above 25 percent indicates significant problems that require remediation (**Figure 3** and **Table 2**).
3. TER as an overall indicator of productivity with the system.

Taking measurements is important in all cases, even when things appear to be going well based on user feedback and practitioner observation. In this study, those with good expertise had low error rates, confirming their skill, but the TER measurements were much lower than one might have guessed when just observing. Recording the actual TER can reveal the need to look further for configuration changes that can enhance performance for these individuals. For those who appear to be struggling, measurements can reveal just how significant that struggle is. Subjects 002, 003, 006, and 009 all experienced scanning errors above 50 percent, with the latter two exceeding 100 percent. In the latter case, this means they were making more than one scanning error for every correct item selection. This is really a crisis situation that demands intervention to reduce frustration and maximize potential.

Carefully configuring a scanning system to match an individual's specific needs seems to yield significant benefits. This is true even if the user appears to be using their system effectively. For example, in subject 004's case, his baseline expertise was high. However, with careful analysis, we were able to identify multiple options for reducing the scan distance in his system and enhancing TER. The best approach is to follow the method individually for each person. However, we observed a few tendencies in this study that tend to apply to everyone. These include—

1. For group organization, such as shown in **Figure 1**, it generally works best to put the letters+prediction group first, and, once selected, to stay on that group until the word is completed.
2. Character prediction typically does not provide performance benefits.
3. Be sure there is enough time for the user to use any prediction features effectively, i.e., to search and select the list on the first pass.

Some of these individuals may be well served by alternatives to single-switch scanning, although that was not addressed in this study. An option for some is the use of eye gaze for those who have both the motor control and perceptual ability to make direct selections using their eyes. We have found very little relevant performance data in the literature, but one study does suggest an advantage of eye gaze over single-switch scanning for those who are able to use either method [23]. However, the scanning matrix in that study used an inefficient alphabetic matrix and had no prediction features. A key in making the decision is a careful assessment of an individual's abilities, including quantitative measurements and trials under appropriate configurations.

Limitations

The primary limitation in this study was the limited set of conditions in the baseline settings. All subjects used text- or letter-based items in their systems, and they used only single-switch automatic scanning in a group-row-column or row-column pattern. While this general setup is common and reflects the actual systems used by these participants in daily life, it does not cover the full range of possibilities available within the scanning input method. For example, people with symbol-based systems did not participate, nor did those using two switches or other single-switch scan patterns such as inverse or step scanning. Future work would be needed to expand our approach to cover other types of scanning configurations. However, similar principles apply, and many of the configuration adjustments work for any type of scanning, such as optimizing item locations for scan distance, identifying and accommodating sources of error, and using language features appropriately.

The measurements for scanning errors and TER were based on two-sentence transcription tasks. This creates a consistent approach that allows for clear error identification and group statistical treatment of the data, but it differs somewhat from subjects' real-world task of generating novel text during spoken or written communication. However, we expect that the effect of revised settings would still be significant even if a text composition task were used in the study; other studies that used both text composition and text copy tasks found that task type did not influence the main effect of text input method [24–25]. Two sentences is a fairly small sample for measuring TER in each session. We chose this length to keep the session duration within an hour or so for each participant.

This required only a few minutes for the fastest participants to complete, but a good solid hour of text entry for those with the slowest rates. Given the factor of 10 difference between the minimum and maximum TER's in this study, a relatively short text sample was necessary to ensure a reasonable time burden on participants.

As in all text entry studies, the exact content of the text is a factor in performance. This is especially true in studies of single-switch scanning, where the item position dramatically affects the time required to select that item. For example, sentences whose letters have a shorter scan distance on average will take less time to enter than those whose letters are farther down in the matrix. To minimize this effect, we matched the scan steps in all sentences used in this study to within 10 percent for letters-only entry across six different letter layouts; i.e., for a given letter layout, the scan steps for the sentences varied by no more than 10 percent. However, sentences could not be matched for equal scan steps in the freeform condition given the varied behavior of different prediction schemes and single-word items across the nine different systems used by our participants. Still, the care taken to match scan steps for letters-only entry was taken as a reasonable assurance that sentence content would not be a large factor in user performance.

Another factor in the results is the duration of the study. Participants had at least 4 wk to get familiar with their revised settings, and we measured their performance with those settings at four roughly equal intervals during that time. All subjects seemed quite acclimated to the new settings by the end of the study, but, as shown in **Table 6**, performance seems to have still been improving by session B4. We have no way of knowing what the results would be with a longer protocol. With a shorter protocol, e.g., one stopping at session B3, the effect of revised settings would still have been significant, yielding about a 60 percent improvement compared with baseline.

An important limitation in putting this approach into practice is that it currently can take a significant amount of time and effort to identify the appropriate revised settings for each individual. For example, to decide whether and how to adjust the position and behavior of any word prediction groups, the research team entered text using a variety of layouts and behaviors on the user's system and counted the scan steps involved in each method. We then typically selected the method that yielded the fewest scan steps. This was perhaps the most time-intensive step in the entire procedure. However, after doing this for sev-

eral individuals, similar conclusions emerged that may apply to others, as noted in the "Questionnaire Results" section. The adjustment of other settings was much more straightforward and time-efficient. For example, establishing the proper scan delay requires only a run through the Compass Switch test, which typically takes less than 5 min.

Taking the measurements themselves can also be time consuming. The Compass Switch and Scan tests can measure switch hit time and scanning errors within just a few minutes, but to measure TER and errors on each individual's own system, we relied on video analysis. The presence of data logging within the AAC system did streamline this process quite a bit, when available, since it provided a timestamped log of each item selected by the user. This allowed us to move through the video more quickly.

Future Work

Key areas for future work include the following:

1. Enhance the adjustment procedure to make it more efficient and practical for real clinical use. This includes identifying areas where the procedure can be streamlined as well as where it can be automated. We plan to work on embedding this procedure into a software tool that guides users through each step of the process and supports it with integrated measurement and simulation tools.
2. Expanding the procedure and tests of its effectiveness to a broader range of individuals and types of scanning systems.

CONCLUSIONS

The results for this approach show great potential for enhancing the TER achieved by single-switch scanning users. By more than doubling their TER, on average, participants gained significant function by using the revised settings identified using our procedure. We look forward to continued work in this area.

ACKNOWLEDGMENTS

REFERENCES

1. Foulds R. Communication rates for nonspeech expression as a function of manual tasks and linguistic constraints. Proceedings of the International Conference on Rehabilitation Engineering; 1980; Toronto, Canada. p. 83–87.

2. Koester HH, Levine SP. Learning and performance of able-bodied individuals using scanning systems with and without word prediction. Assist Technol. 1994;6(1):42–53. [PMID:10147209] http://dx.doi.org/10.1080/10400435.1994.10132226

3. Lesher G, Moulton BJ, Higginbotham DJ, Alsofrom B. Acquisition of scanning skills: The use of an adaptive scanning delay algorithm across four scanning displays. Proceedings of the 25th Annual Conference on Rehabilitation Engineering (RESNA); 2002 Jun; Atlanta, GA. p. 75–77.

4. MacKenzie IS, Felzer T. SAK: Scanning ambiguous keyboard for efficient one-key text entry. ACM Trans Comput Hum Interact. 2010;17(11):11–39.

5. Mankowski R, Simpson RC, Koester HH. Validating a model of row-column scanning. Disabil Rehabil Assist Technol. 2013;8(4):321–29. [PMID:23078232] http://dx.doi.org/10.3109/17483107.2012.721159

6. Bhattacharya S, Samanta D, Basu A. Performance models for automatic evaluation of virtual scanning keyboards. IEEE Trans Neural Syst Rehabil Eng. 2008;16(5):510–19. [PMID:18990655] http://dx.doi.org/10.1109/TNSRE.2008.2003377

7. Cronk S, Wang W. Investigating relationships between user performance and scan delays in aids that scan. Proceedings of the 25th Annual Conference of the Rehabilitation Engineering and Assistive Technology Society of North America; 2002; Minneapolis, MN. p. 129–31.

8. de Lange C. Saving Stephen's voice. New Sci. 2012;(2846):27. http://dx.doi.org/10.1016/S0262-4079(12)60042-6

9. Fager S, Beukelman DR, Fried-Oken M, Jakobs T, Baker J. Access interface strategies. Assist Technol. 2011;24(1):25–33. [PMID:22590797] http://dx.doi.org/10.1080/10400435.2011.648712

10. Lesher GW, Moulton BJ, Higginbotham DJ. Techniques for augmenting scanning communication. Augment Altern Commun. 1998;14(2):81–101. http://dx.doi.org/10.1080/07434619812331278236

11. Horstmann HM. Quantitative modeling in augmentative communication—a case study. Proceedings of the 13th Annual Conference on Rehabilitation Engineering (RESNA); 1990 Jun; Washington, DC.

12. Damper RI. Text composition by the physically disabled: A rate prediction model for scanning input. Appl Ergon. 1984;15(4):289–96. [PMID:15676527]

13. Anson DK. Alternative computer access: A guide to selection. Philadelphia (PA): F. A. Davis; 1997.

14. Angelo J. Comparison of three computer scanning modes as an interface method for persons with cerebral palsy. Am J Occup Ther. 1992;46(3):217–22. [PMID:1532689] http://dx.doi.org/10.5014/ajot.46.3.217

15. Simpson R, Koester H, LoPresti E. Selecting an appropriate scan rate: The ".65 rule." Assist Technol. 2007;19(2):51–58, quiz 59–60. [PMID:17727073] http://dx.doi.org/10.1080/10400435.2007.10131865

16. Simpson RC, Mankowski R, Koester HH. Modeling one-switch row-column scanning with errors and error correction methods. Open Rehabil J. 2011;1–12.

17. Smith JM, Simpson RC. Analyzing performance with computer access technology using unconstrained text entry protocol. J Rehabil Res Dev. 2009;46(8):1059–68. [PMID:20157863] http://dx.doi.org/10.1682/JRRD.2008.09.0134

18. Nielsen J. Usability engineering. Boston (MA): Academic Press; 1993.

19. Goodenough-Trepagnier C, Galdieri B, Rosen MJ, Baker E. Language anticipation with a computer based scanning communication aid. Proceedings of the IEEE Computer Society Workshop on Computing to Aid the Handicapped; 1984 Nov; Charlottesville, VA. p. 11–15.

20. Simpson R, Koester HH, LoPresti E. Evaluation of an adaptive row/column scanning system. Technol Disabil. 2007;18:1–12.

21. Simpson RC, Koester HH. Adaptive one-switch row-column scanning. IEEE Trans Rehabil Eng. 1999;7(4):464–73. [PMID:10609634] http://dx.doi.org/10.1109/86.808950

22. Roark B, deVilliers J, Gibbons C, Fried-Oken M. Scanning methods and language modeling for binary switch typing.

Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies; 2010 Jun 5; Los Angeles, CA. p. 28–36.

23. Gibson C, Beneteau E. Functional performance using eye control and single-switch scanning by people with ALS. In: Perspectives on augmentative and alternative communication. Rockville (MD): American Speech-Language-Hearing Association; 2010. p. 64–69.

24. Karat CM, Halverson C, Horn D, Karat J. Patterns of entry and correction in large vocabulary continuous speech recognition systems. Proceedings of the Conference on Human Factors in Computing Systems; 1999 May 15–20; Pittsburgh, PA. p. 568–75.

25. Cox AL, Cairns PA, Walton A, Lee S. Tlk or txt? Using voice input for SMS composition. Pers Ubiquitous Com-

put. 2008;12:567–88.
http://dx.doi.org/10.1007/s00779-007-0178-8